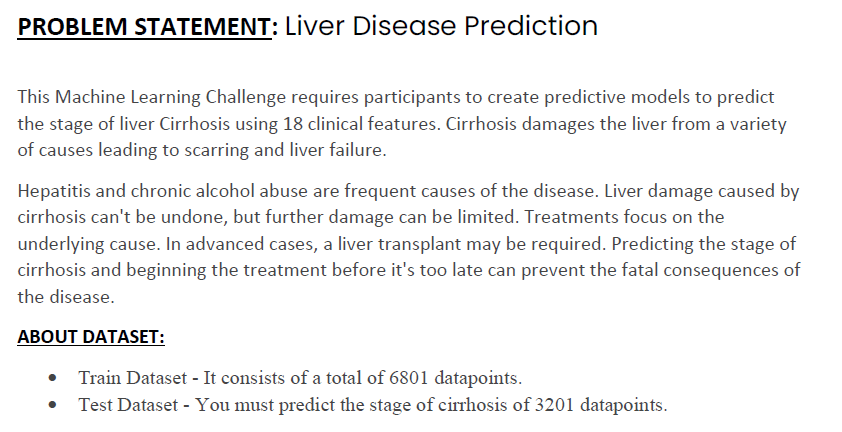
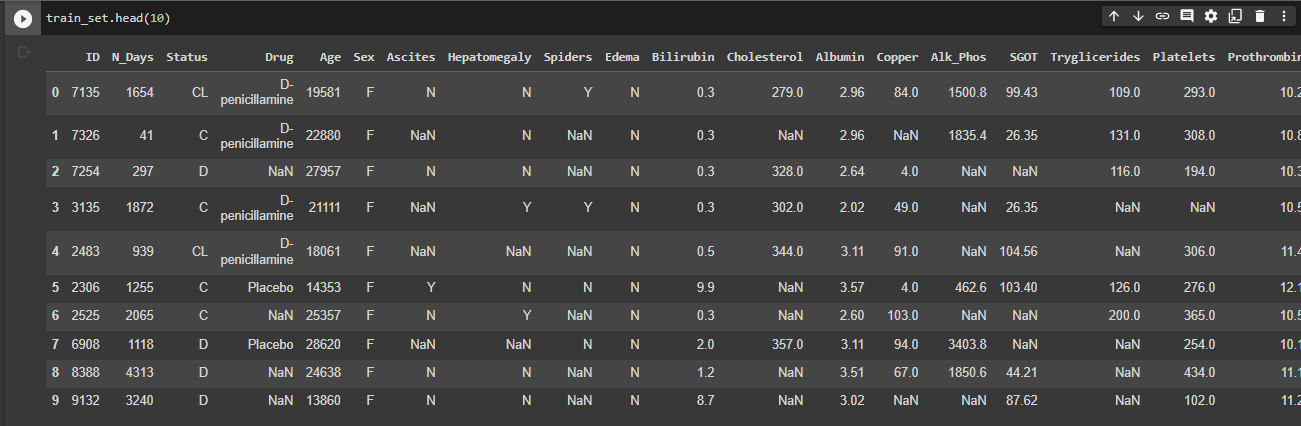
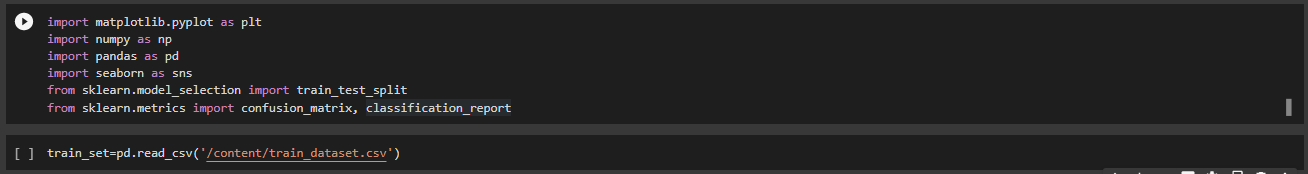
**Predicting Liver Disease Using Machine Learning Model**

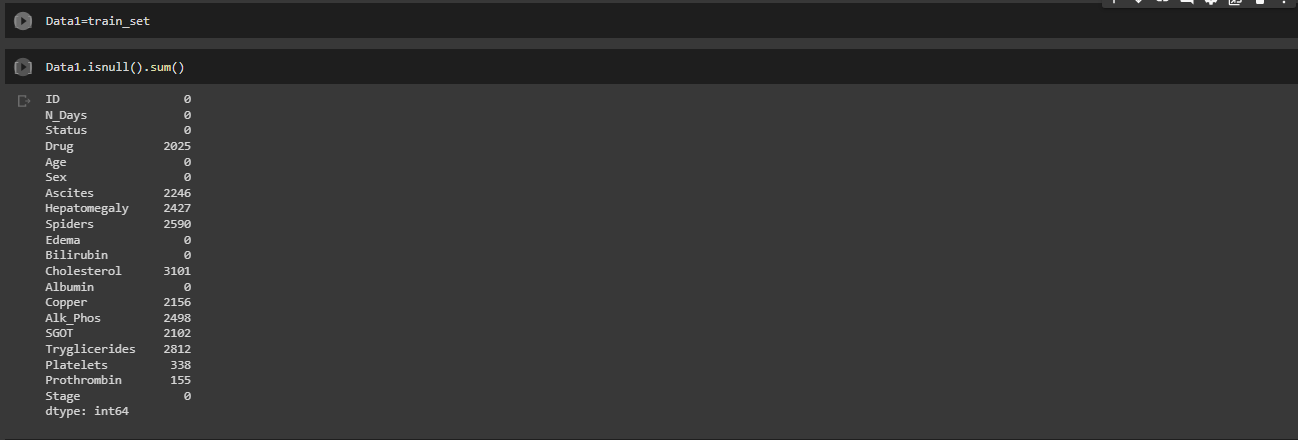
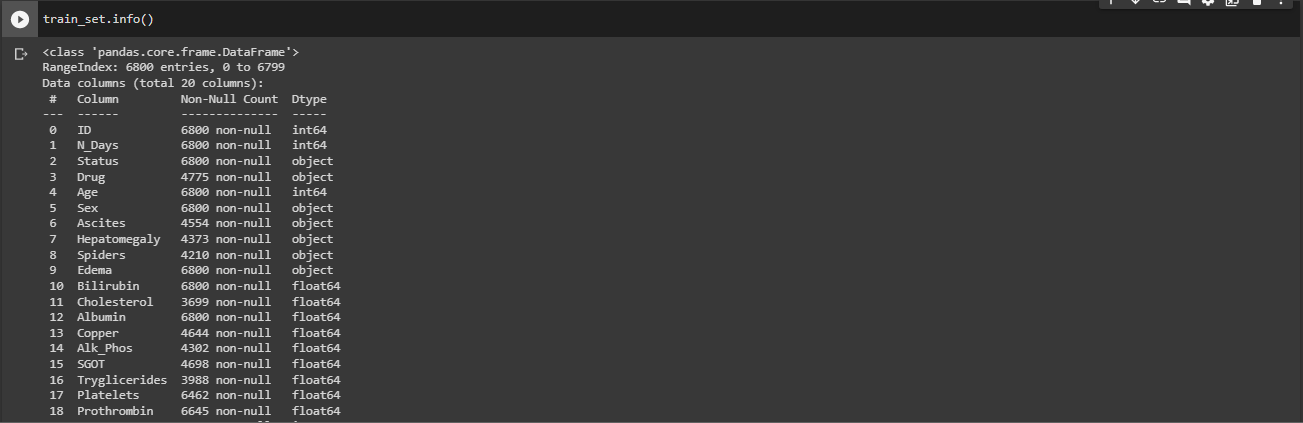


In this dataset, we have separate data set for training the ML Model and Testing the Model. So, lets Approach them Separately

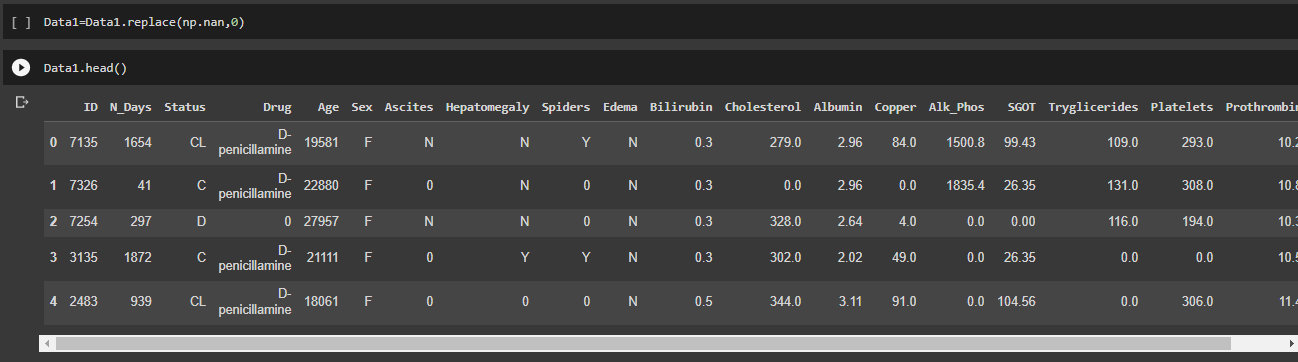
Training Data Analysis

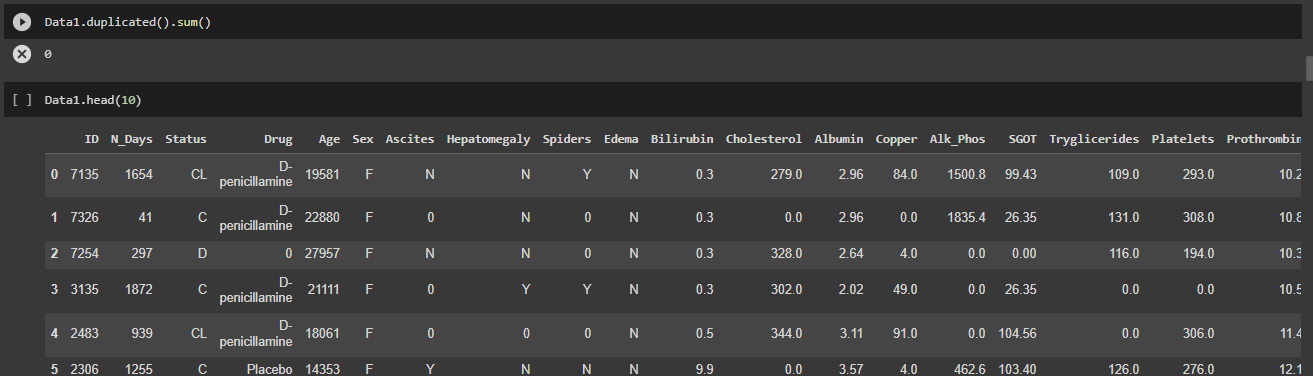
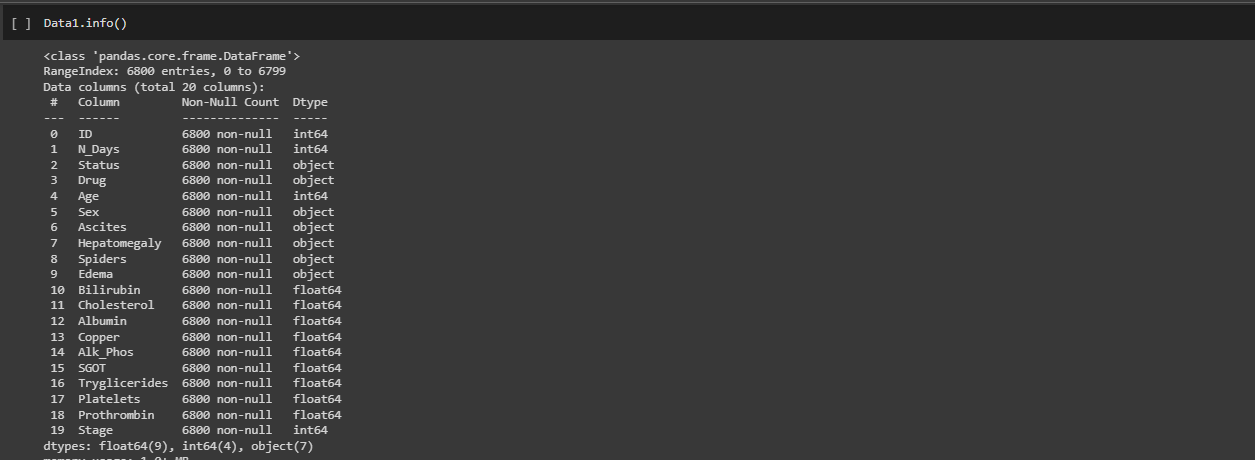
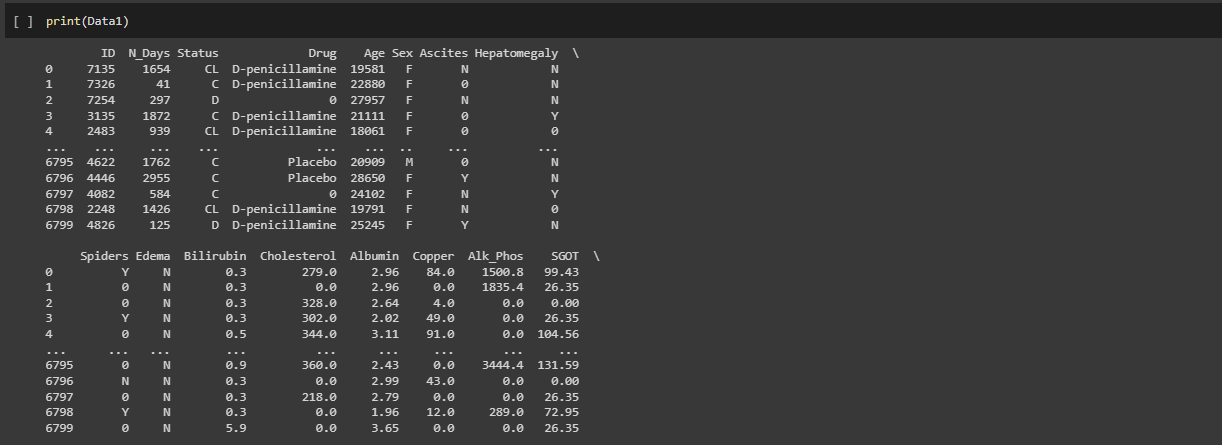
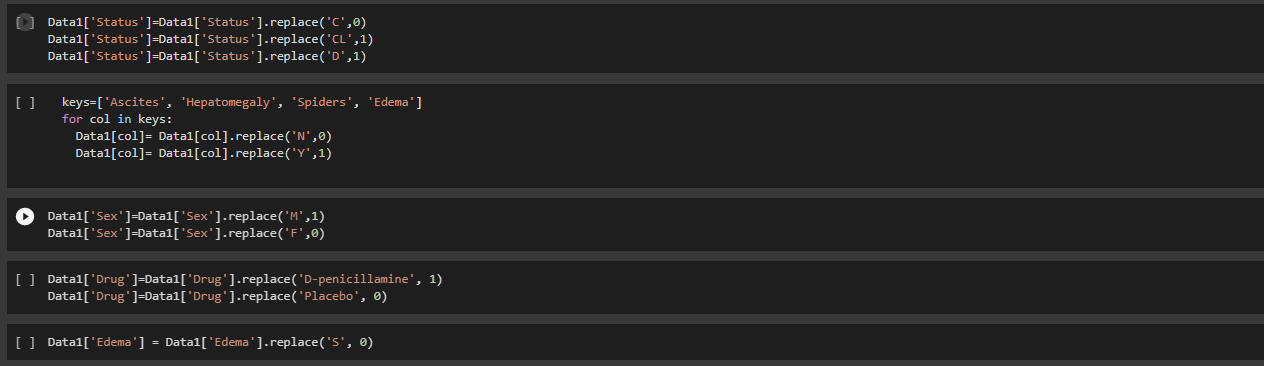
**Step 1**: Gathering the data set, Loading and Reading the Data set.

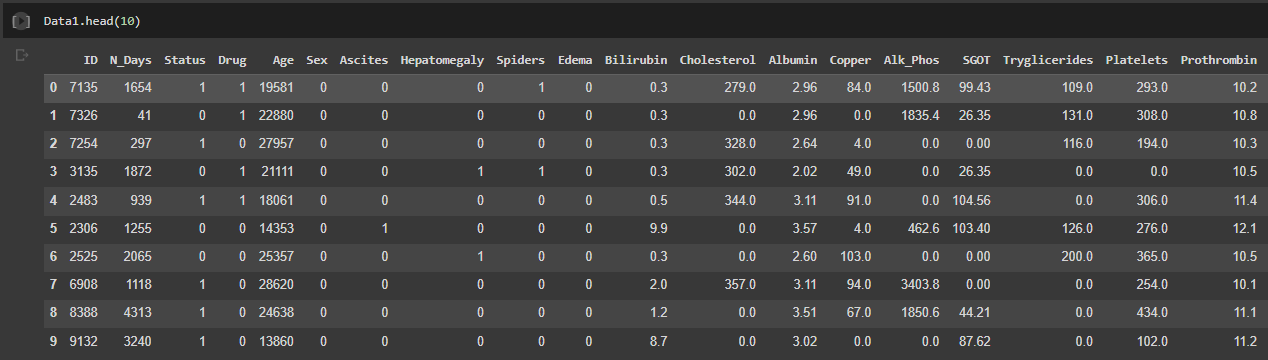
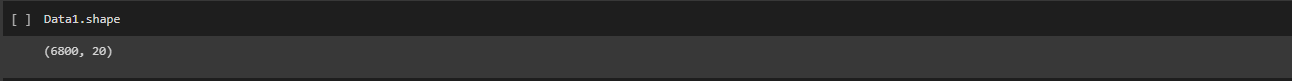
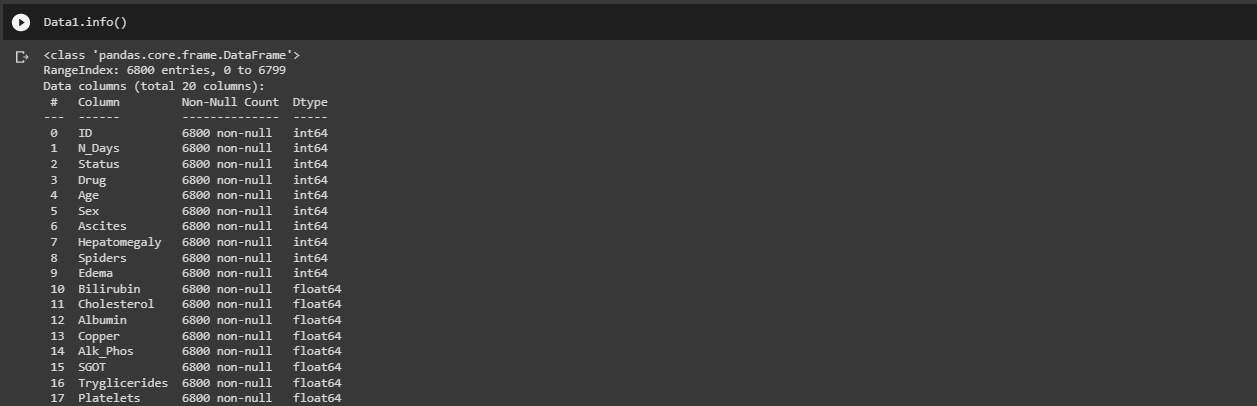




**Step 2**: Performing Hot Encoding/Data cleaning to convert Categorical Features into Numerical Values.



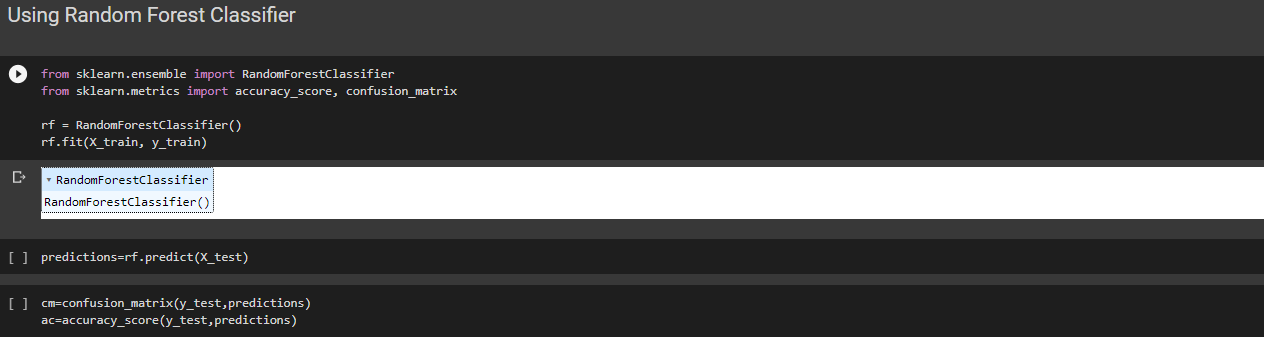
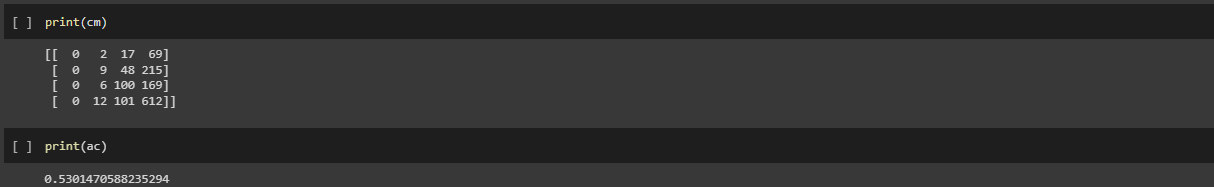
**Step 3:** Dividing the data into Dependent(y) and Independent(x) variables.



**Step 4**: Splitting the data into Training and Testing set.

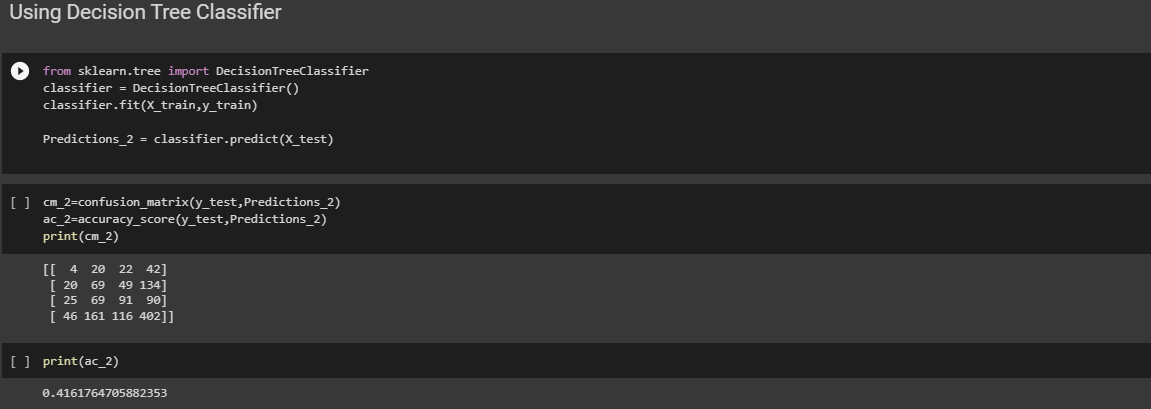


**Step 5a:** Creating Machine Learning Model using **Random Forest Classifier.**

So we can obsereve that the accuracy of prediction is 53% in case of Random Forest Classifier. Lets try if Decision Tree Classifier gives the better results.

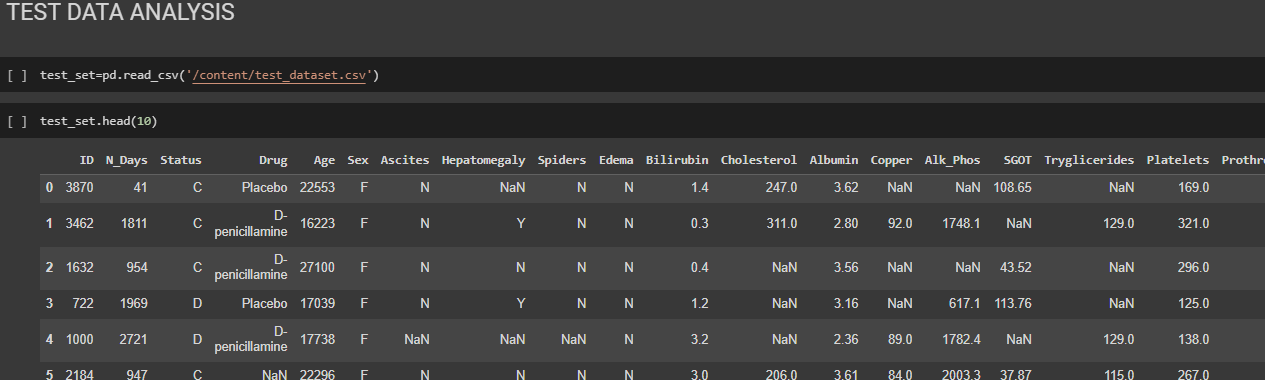
**Step 5b**: Creating Machine Learning Model using **Decision Tree Classifier.**



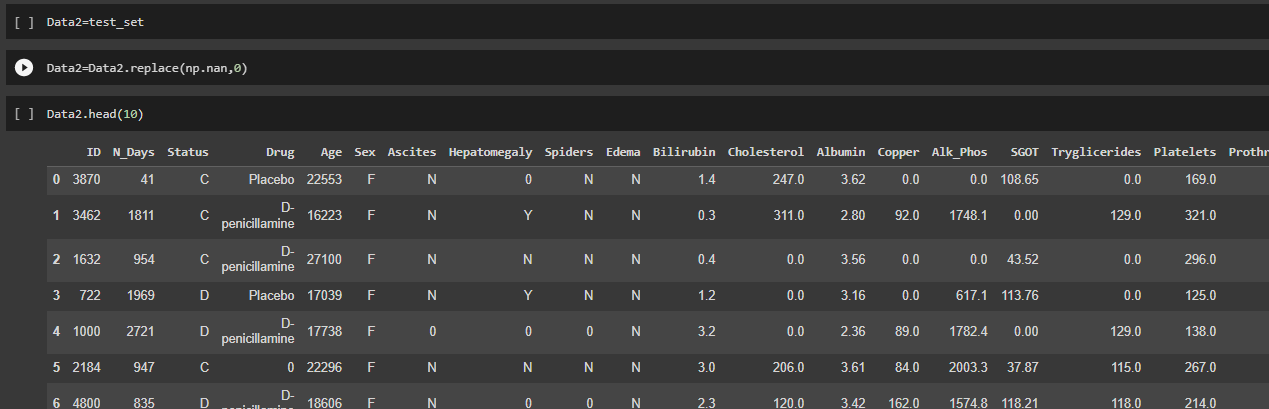
So the accuracy of the prediction is 41.61% in case of Decision Tree Classifier. So comparatively Random Forest Classifier is giving the Better Results.

Testing Data Analysis

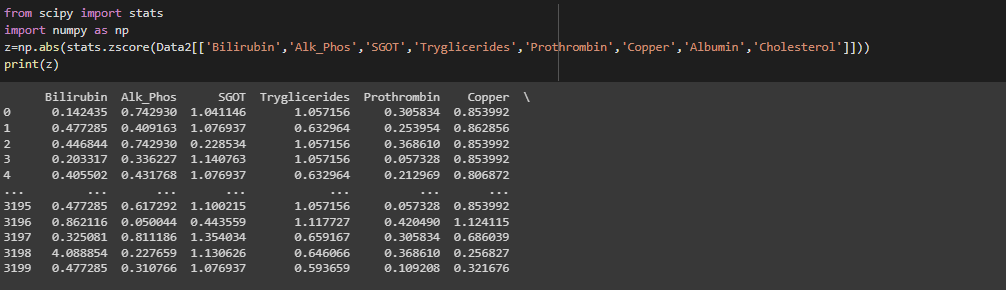
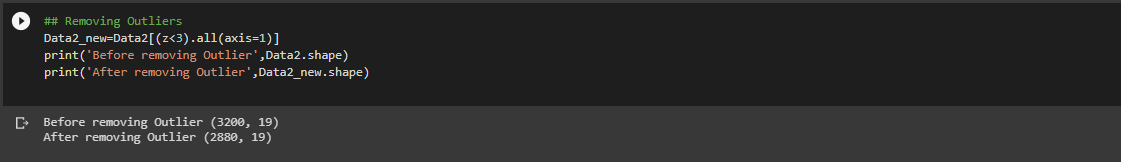
**Step 1:** Gathering the data set, Loading and Reading the Data set.



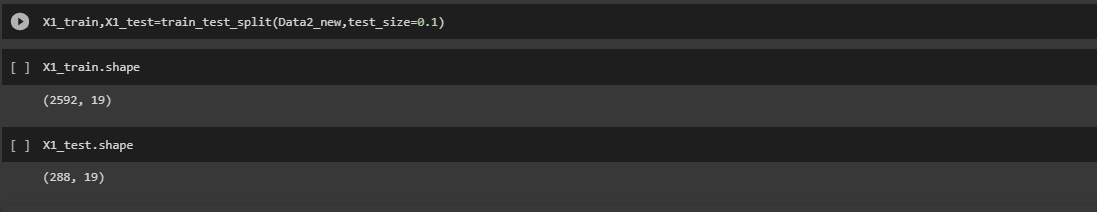
**Step 2:** Performing Hot Encoding/Data cleaning to convert Categorical Features into Numerical Values.

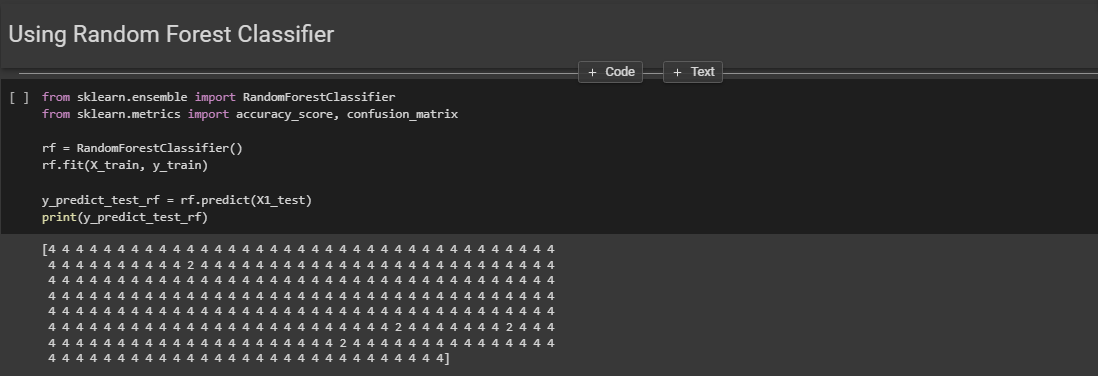
Since these the too many rows of data for Testing Set, there are chances that outliers might affect the prediction. So let’s remove the outliers.

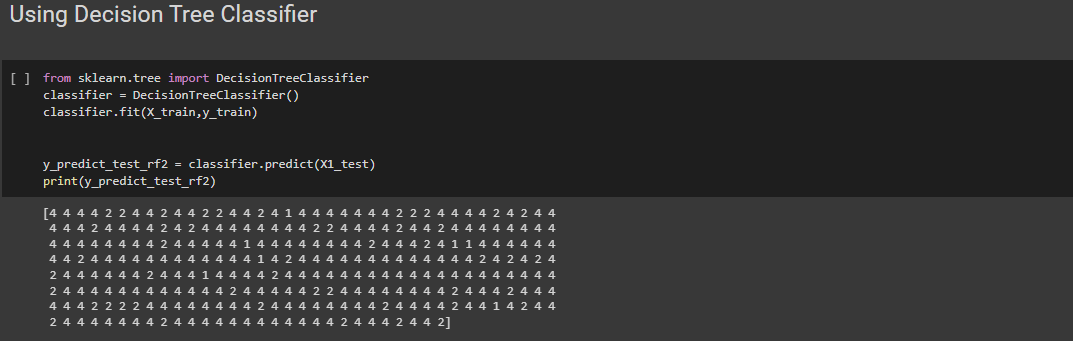
**Step 3:** Splitting the data into Training and Testing set.



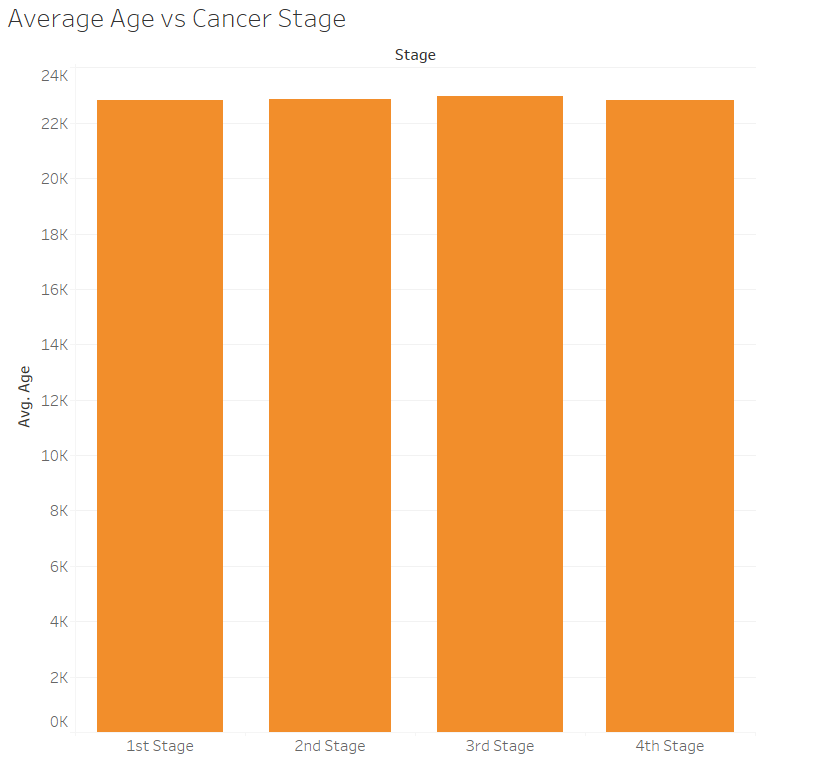
**Step 4a:** Predicting the Cancer stage using **Random Forest Classifier.**



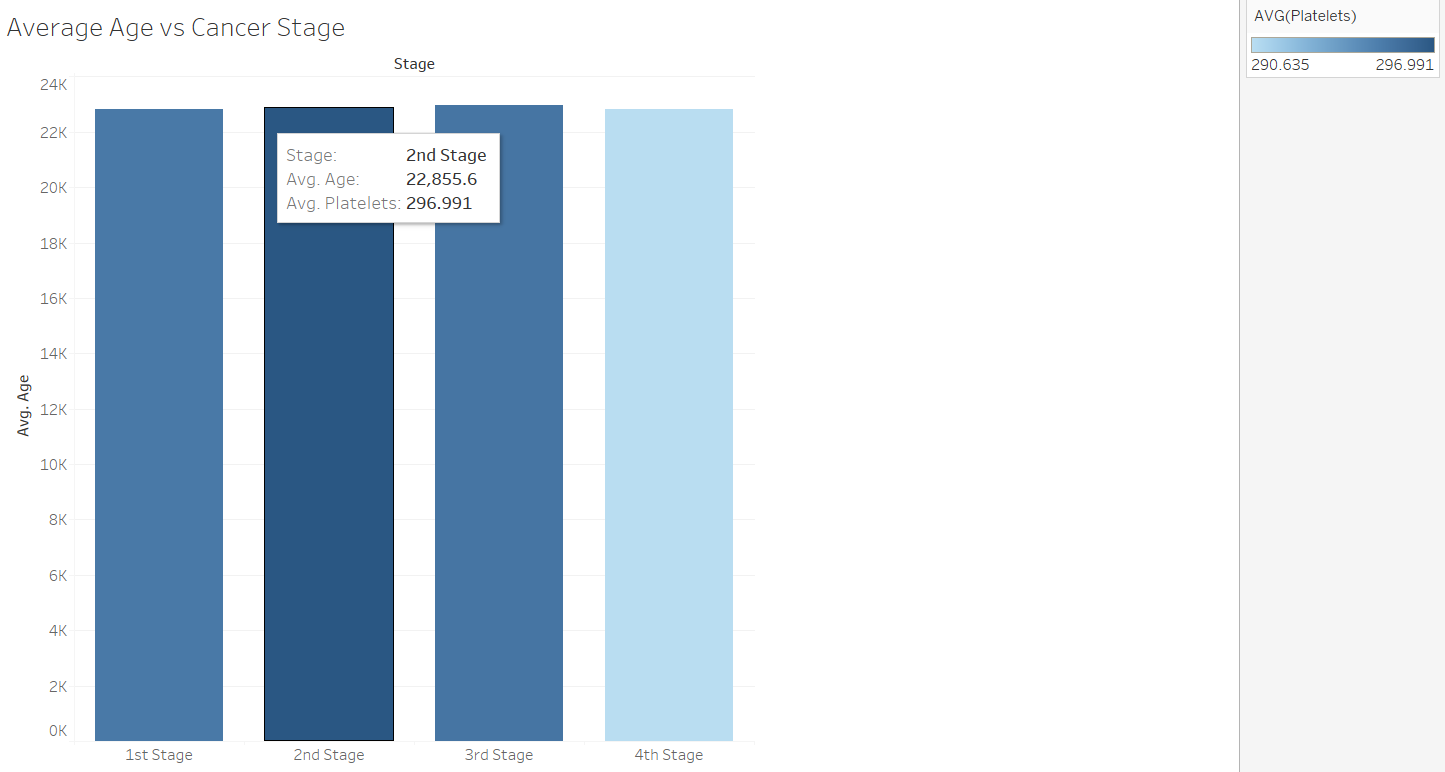
**Step 4b:** Predicting the Cancer stage using **Decision Tree Classifier.**



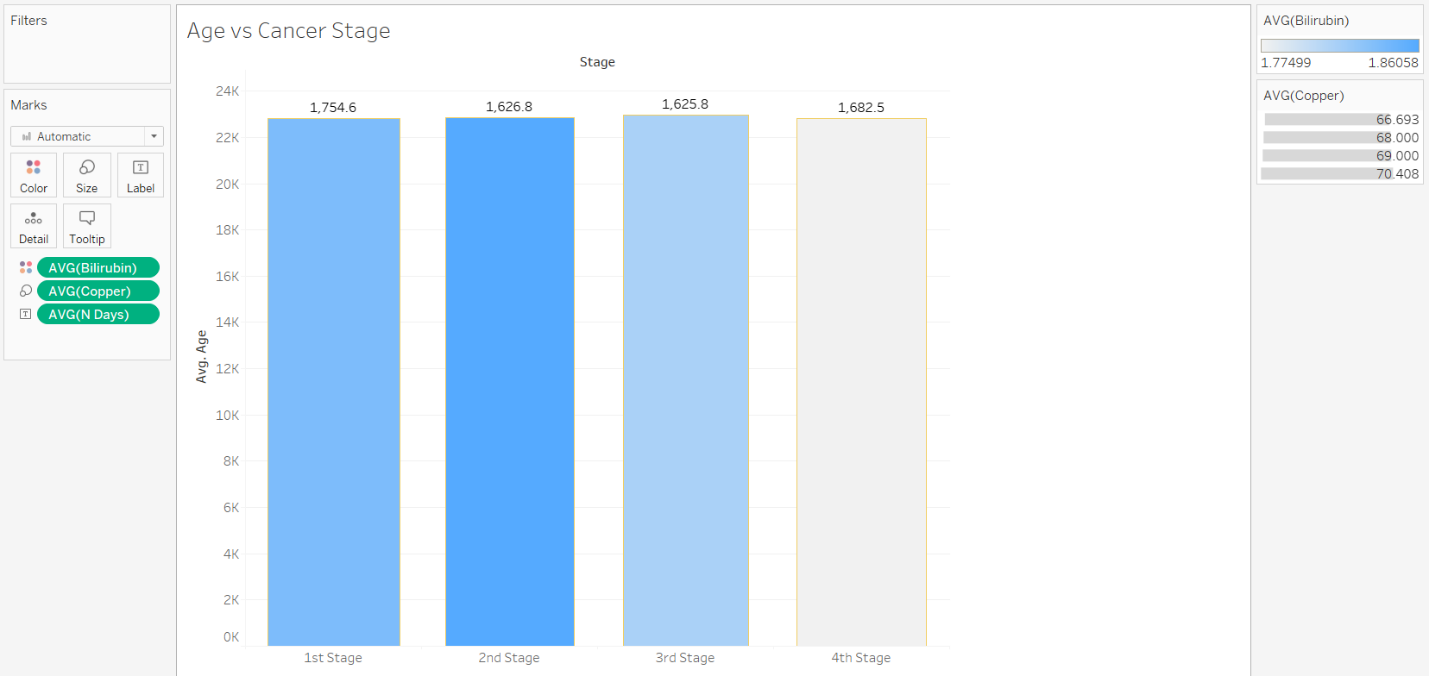
**Visualization of Data:**



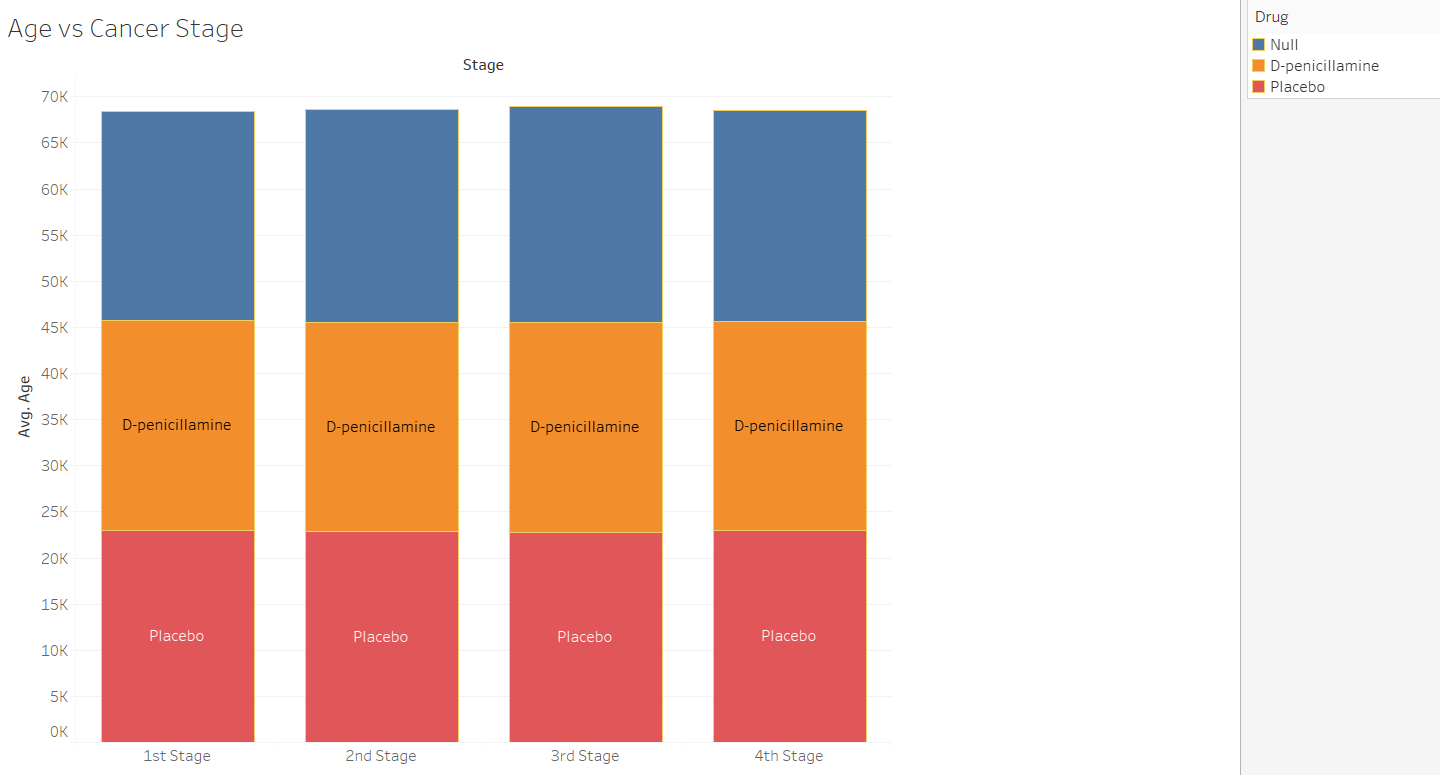
From this graph of Age(in days) vs Cancer stages, we can observe that level of cholestrol is independent on the stages of cancer. Thickness of Bar graph is almost same in all the stages so its Independent Parameter.



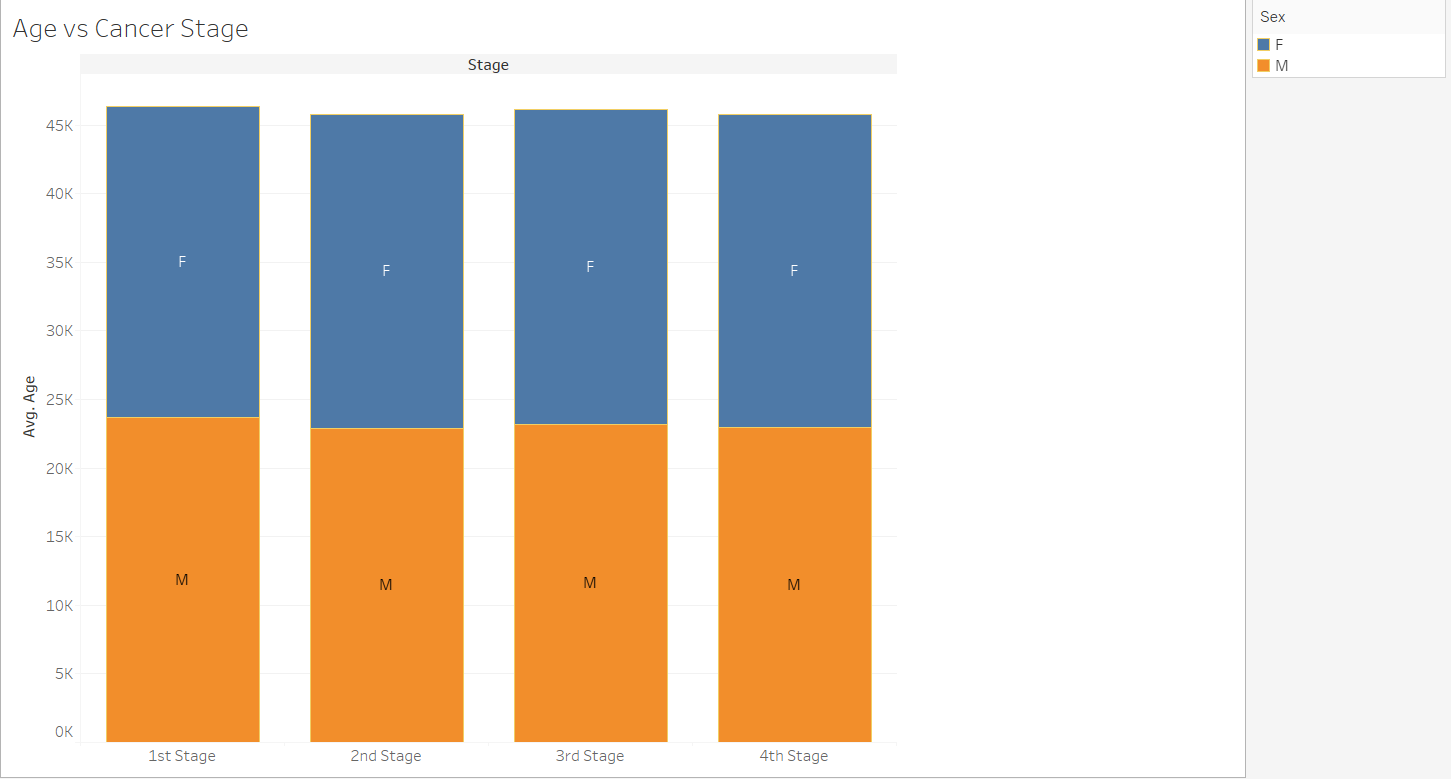
From the above graph we can observe that as the patient gradually moves from 2nd to 4th Stage, the platelets counts gradually decreases.



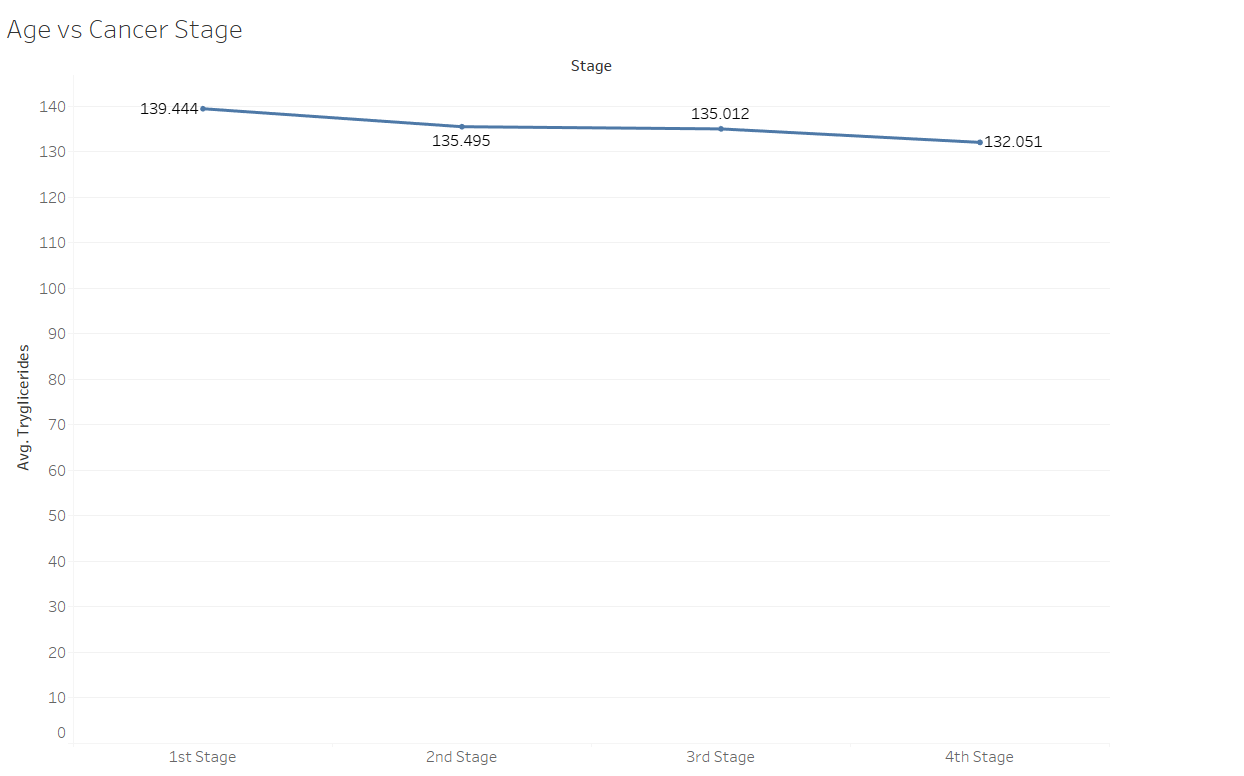
From the above graph we can observe that Number of days spent in 1st stage is more compared to 2nd, 3rd and 4th Stage. Bilirubin content is found to decrease significantly from 2nd to 4th Stage. There is no change in copper content in any of the Stages.



This graph signifies that there is no much change in the drug used in Different Stages of cancer.



The ratio of Male to Female in different Stages is almost same.



From the above graph we can see that the level of Tryglicerides gragually reduces as the stage proceeds from 1st to 4th .